浙江大学
ZHEJIANG UNIVERSITY
Mohamed bin Zayed University of Artificial Intelligence
埃默里大学
EMORY UNIVERSITY
EMORY
ICLR | 2025
Thirteenth International Conference on Learning Representations

# Causal Graph Transformer for Treatment Effect Estimation Under Unknown Interference

Anpeng Wu[1,2], Haiyi Qiu[1], Zhengming Chen[2,3], Zijian Li[2], Ruoxuan Xiong[4], Fei Wu[1*], Kun Zhang[2,5*]

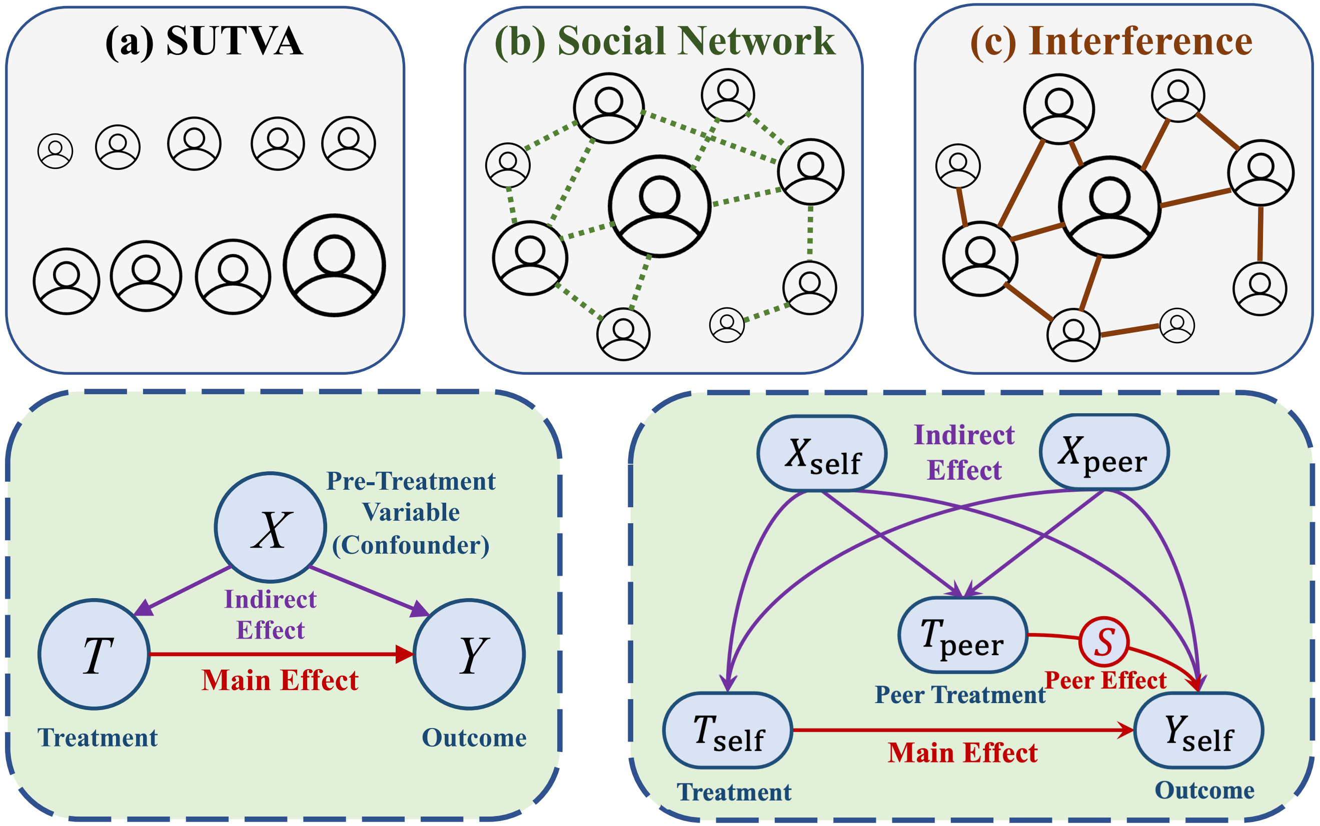
1 Department of Computer Science and Technology, Zhejiang University, Hangzhou, China
2 Department of Machine Learning, MBZUAI, Abu Dhabi, UAE
3 Guangdong University of Technology, Guangzhou, China
4 Department of Quantitative Theory and Methods, Emory University, Atlanta, USA
5 Department of Philosophy, Carnegie Mellon University, Pittsburgh, USA
anpwu@zju.edu.cn
*Corresponding author

广东工业大学
Guangdong University of Technology
Carnegie Mellon University

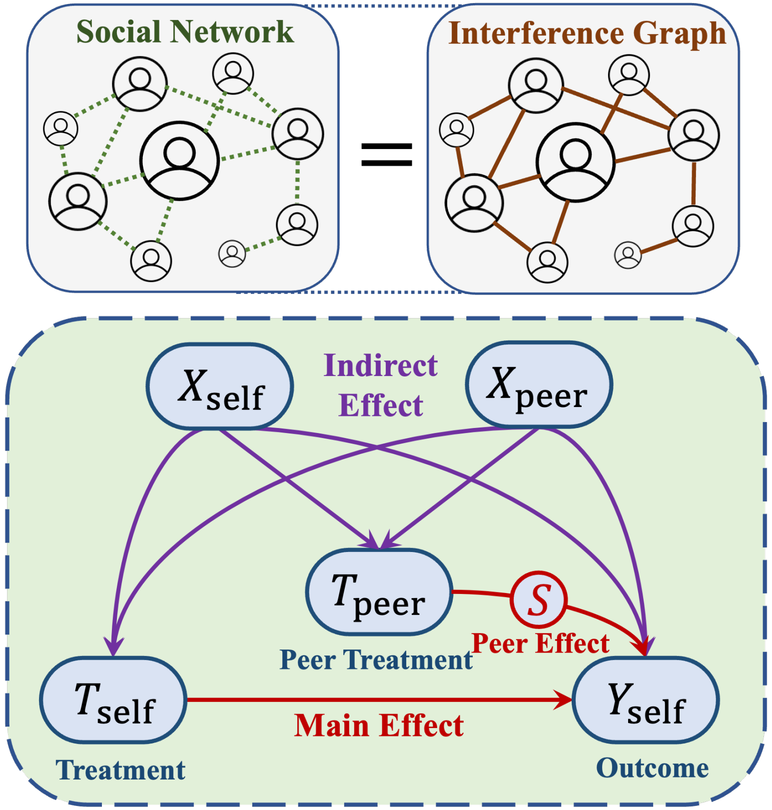
# Treatment Effect Estimation Under Unknown Interference.

**Stable Unit Treatment Value Assumption (SUTVA)** has two main components:

1. **No Interference**: Each individual's potential outcome is only influenced by their own treatment status and not by the treatment status of others.
2. **No Hidden Versions of Treatment**: Each treatment has a single, consistent version without variation.



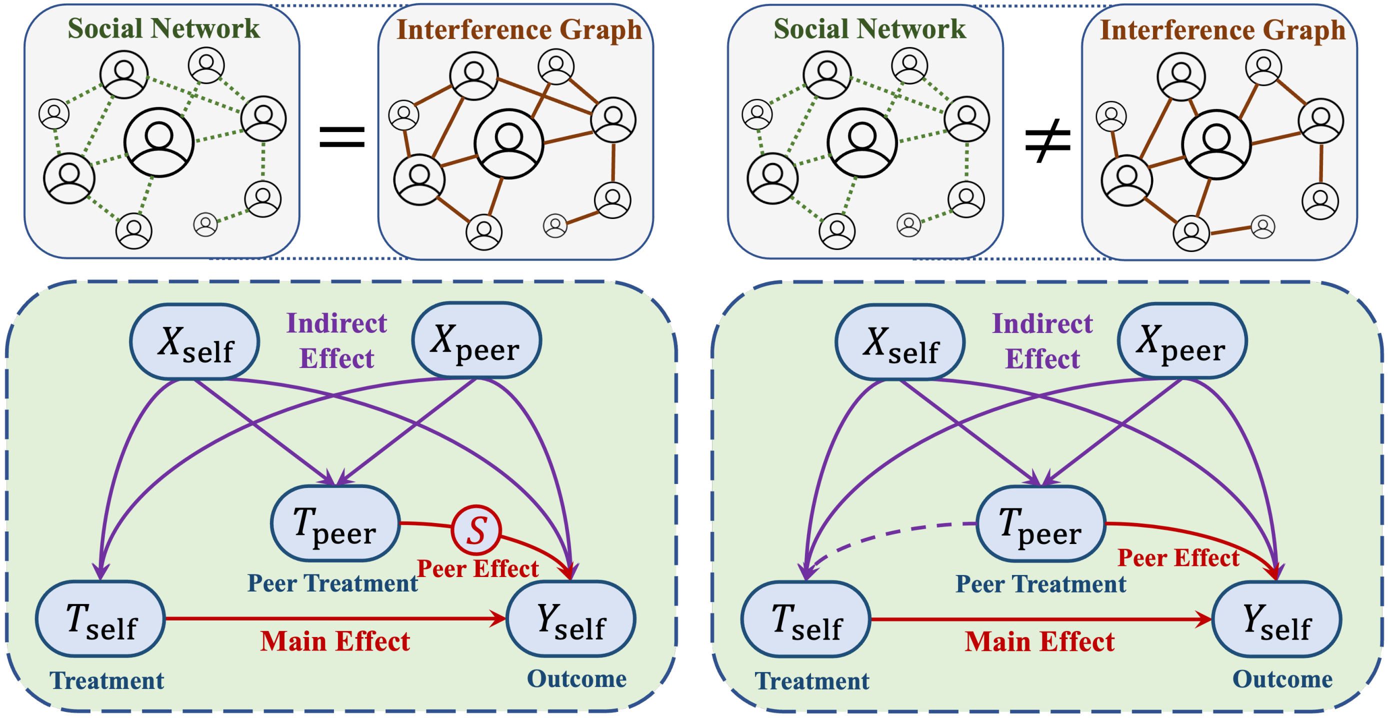

(d) SUTVA Setting (e) Network Setting



# Treatment Effect Estimation Under Unknown Interference.

**Stable Unit Treatment Value Assumption (SUTVA)** has two main components:

1. **No Interference**: Each individual's potential outcome is only influenced by their own treatment status and not by the treatment status of others.
2. **No Hidden Versions of Treatment**: Each treatment has a single, consistent version without variation.




**(a) Traditional Network Setting** **(b) General Network Setting**

## Different from Traditional Network Settings

1. The interference graph is not the same as the social network; the networked interference structure is unknown.

2. The structural function describing the effect of peer treatments ($T_{peer}$) on self-outcomes ($Y_{self}$), also known as the summary/aggregation function of exposure mapping, is unknown.

3. In the general network setting, there may be potential causal effects between peer treatments ($T_{peer}$) and self-treatments ($T_{self}$).

# Representative Algorithms on Observational Networked Data.

| | Settings | | Effects | | | Prior | | Sub-Modules | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | Interference | Unmeasured | Main | Peer | Total | Graph | Summary | Reweighting | Representation | Attention |
| CNE (Veitch et al., 2019) | | P | ✓ | | | ✓ | | ✓ | | |
| NetDeconf (Guo et al., 2020) | | P | ✓ | | | ✓ | | | ✓ | |
| DRLearner (Leung & Loupos, 2022) | ✓ | | ✓ | ✓ | ✓ | ✓ | | ✓ | | |
| SPNet (Huang et al., 2023) | ✓ | P | ✓ | | | ✓ | | ✓ | ✓ | ✓ |
| Net-TMLE (Ogburn et al., 2024) | ✓ | | ✓ | | | ✓ | ✓ | ✓ | | |
| GDML (Khatami et al., 2024) | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | |
| G-HSIC (Ma & Tresp, 2021) | ✓ | | ✓ | | | ✓ | ✓ | | ✓ | |
| RRNet (Cai et al., 2023) | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| NetEst (Jiang & Sun, 2022) | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | |
| Uncertainty (Bhattacharya et al., 2020) | ✓ | | | | ✓ | | | | | |
| UNITE (Lin et al., 2024) | ✓ | | ✓ | | | | | | ✓ | |
| CauGramer (Ours) | ✓ | ✓ | ✓ | ✓ | ✓ | | | | ✓ | ✓ |

- **Without interference effects:** CNE and NetDeconf algorithms use reweighting and balanced representation learning;
- **Known interference graph:** DRLearner, SPNet, Net-TMLE, GDML, G-HSIC, RRNet, NetEst;
- **Unknown interference:**
  - Uncertainty algorithm proposes a method integrating structure learning and causal inference to estimate the Population Average Overall Effect (PAOE) under network uncertainty and partial interference.
  - UNITE algorithm uses a Graph Structure Learner to infer the hidden interference structure by constructing a complete graph and imposing L0-norm regularization to identify significant connections.

# Parameters of Interest.

**Definition 1** (Individual Main Effects (IME)). *IME denotes the effects of self-treatment, i.e.,* $\tau_{\text{IME}}(\boldsymbol{x}_i) = y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 1, \mathbf{0}_{\mathcal{P}}) - y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 0, \mathbf{0}_{\mathcal{P}})$.

**Definition 2** (Individual Peer Effects (IPE)). *IPE denotes the effects of peers' treatments, i.e.,* $\tau_{\text{IPE}}(\boldsymbol{x}_i, \boldsymbol{t}_{\mathcal{P}}) = y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 0, \boldsymbol{t}_{\mathcal{P}}) - y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 0, \mathbf{0}_{\mathcal{P}})$ *for any* $\boldsymbol{t}_{\mathcal{P}} \in \mathcal{T}^{|\mathcal{P}_i|}$.

**Definition 3** (Individual Total Effects (ITE)). *ITE denotes the combination of main and peer effects, i.e.,* $\tau_{\text{ITE}}(\boldsymbol{x}_i, \boldsymbol{t}_{\mathcal{P}}) = y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 1, \boldsymbol{t}_{\mathcal{P}}) - y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 0, \mathbf{0}_{\mathcal{P}})$ *for any* $\boldsymbol{t}_{\mathcal{P}} \in \mathcal{T}^{|\mathcal{P}_i|}$.

# Identification Assumptions.

To precisely estimate the three treatment effects, i.e., ME, PE, and TE, we first discuss causal identification under the standard causal assumptions on networked data (Jiang & Sun, 2022).

**Assumption 1** (Positivity). *The probability of a unit with their peers to receive any treatment pair* $(t, \boldsymbol{t}_{\mathcal{P}})$ *is always positive, i.e.,* $0 < \mathbb{P}(t_i = t, \boldsymbol{t}_{\mathcal{P}_i} = \boldsymbol{t}_{\mathcal{P}} \mid \boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}) < 1$ *for any* $\boldsymbol{x}_i$.

**Assumption 2** (Consistency). *The potential outcome is the same as the observed outcome under the same self-treatment and peer-treatments, i.e.,* $y_i = y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, t_i, \boldsymbol{t}_{\mathcal{P}_i})$ *for treatment pair* $(t_i, \boldsymbol{t}_{\mathcal{P}_i})$.

**Assumption 3** (Unconfoundedness). *The self-treatment and peer-treatments are independent of the potential outcome given self and peer' features, i.e.,* $y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, t, \boldsymbol{t}_{\mathcal{P}}) \perp\!\!\!\perp (t, \boldsymbol{t}_{\mathcal{P}}) \mid (\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i})$.

# Identification Theorems.

**Theorem 1** (Causal Identification). *Given these assumptions, while the Stable Unit Treatment Value Assumption (SUTVA) does not hold under networked interference, the treatment effects are identified as long as we can control the confounders $\boldsymbol{x}_i$ and the peers $\boldsymbol{x}_{\mathcal{P}_i}$.*

*Proof.* As shown in Figure 1, when we consider the (unknown) peer interference graph, all common causes $(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i})$ of treatment pair $(t_i, \boldsymbol{t}_{\mathcal{P}_i})$ and outcomes $y_i$ have been discovered. Thus, we have:

$$\begin{aligned}\tau_{\text{ITE}}(\boldsymbol{x}_i, \boldsymbol{t}_{\mathcal{P}}) &= \mathbb{E}[y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 1, \boldsymbol{t}_{\mathcal{P}}) \mid \boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}] - \mathbb{E}[y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 0, \boldsymbol{0}_{\mathcal{P}}) \mid \boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}] \\ &= \mathbb{E}[y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 1, \boldsymbol{t}_{\mathcal{P}}) \mid \boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, t, \boldsymbol{t}_{\mathcal{P}}] - \mathbb{E}[y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 0, \boldsymbol{0}_{\mathcal{P}}) \mid \boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 0, \boldsymbol{0}_{\mathcal{P}}] \quad (1) \\ &= \mathbb{E}[y_{1,\boldsymbol{t}_{\mathcal{P}}} \mid \boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, t, \boldsymbol{t}_{\mathcal{P}}] - \mathbb{E}[y_{0,\boldsymbol{0}_{\mathcal{P}}} \mid \boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 0, \boldsymbol{0}_{\mathcal{P}}], \quad (2)\end{aligned}$$

where $y_{t,\boldsymbol{t}_{\mathcal{P}}}$ is the observed outcome when the unit and its peers have features $\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}$ and receive the treatment pair $(t, \boldsymbol{t}_{\mathcal{P}})$. Eq. (1) holds under the Uncounfoundedness, i.e., $y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, t, \boldsymbol{t}_{\mathcal{P}}) \perp\!\!\!\perp (t, \boldsymbol{t}_{\mathcal{P}}) \mid (\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i})$. Eq. (2) holds under the Consistency Assumption, i.e., $y_{t,\boldsymbol{t}_{\mathcal{P}}} = y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, t, \boldsymbol{t}_{\mathcal{P}})$. Theorem 1 holds for any treatment pair $(t, \boldsymbol{t}_{\mathcal{P}})$ under Positivity Assumption, i.e., $\mathbb{P}(t_i = t, \boldsymbol{t}_{\mathcal{P}_i} = \boldsymbol{t}_{\mathcal{P}}) > 0$. □

However, since the interference graph is unknown, we cannot directly model the expectation function $\mathbb{E}[y_{t,\boldsymbol{t}_{\mathcal{P}}} \mid \boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, t, \boldsymbol{t}_{\mathcal{P}}]$. Nevertheless, under the Unconfoundednes assumption, we know that $y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, t, \boldsymbol{t}_{\mathcal{P}}) \perp \{t_i, \boldsymbol{t}_{\mathcal{P}_i}\} \mid \{\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}\}$. Therefore, if we control for all observed confounders $\{\boldsymbol{x}_i\}_{i=1}^n$, we can infer to $y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, t, \boldsymbol{t}_{\mathcal{P}}) \perp \{t_i, \boldsymbol{t}_{\mathcal{P}_i}\} \mid \{\boldsymbol{x}_i\}_{i=1}^n$. The interaction information in the interference graph is embedded within the node and network features. To capture this, we propose using two functions, $g_x$ and $g_t$, to represent peer feature and peer treatment information, respectively.

# Identification Theorems.

**Theorem 1** (Causal Identification). *Given these assumptions, while the Stable Unit Treatment Value Assumption (SUTVA) does not hold under networked interference, the treatment effects are identified as long as we can control the confounders $\boldsymbol{x}_i$ and the peers $\boldsymbol{x}_{\mathcal{P}_i}$.*

*Proof.* As shown in Figure 1, when we consider the (unknown) peer interference graph, all common causes $(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i})$ of treatment pair $(t_i, \boldsymbol{t}_{\mathcal{P}_i})$ and outcomes $y_i$ have been discovered. Thus, we have:

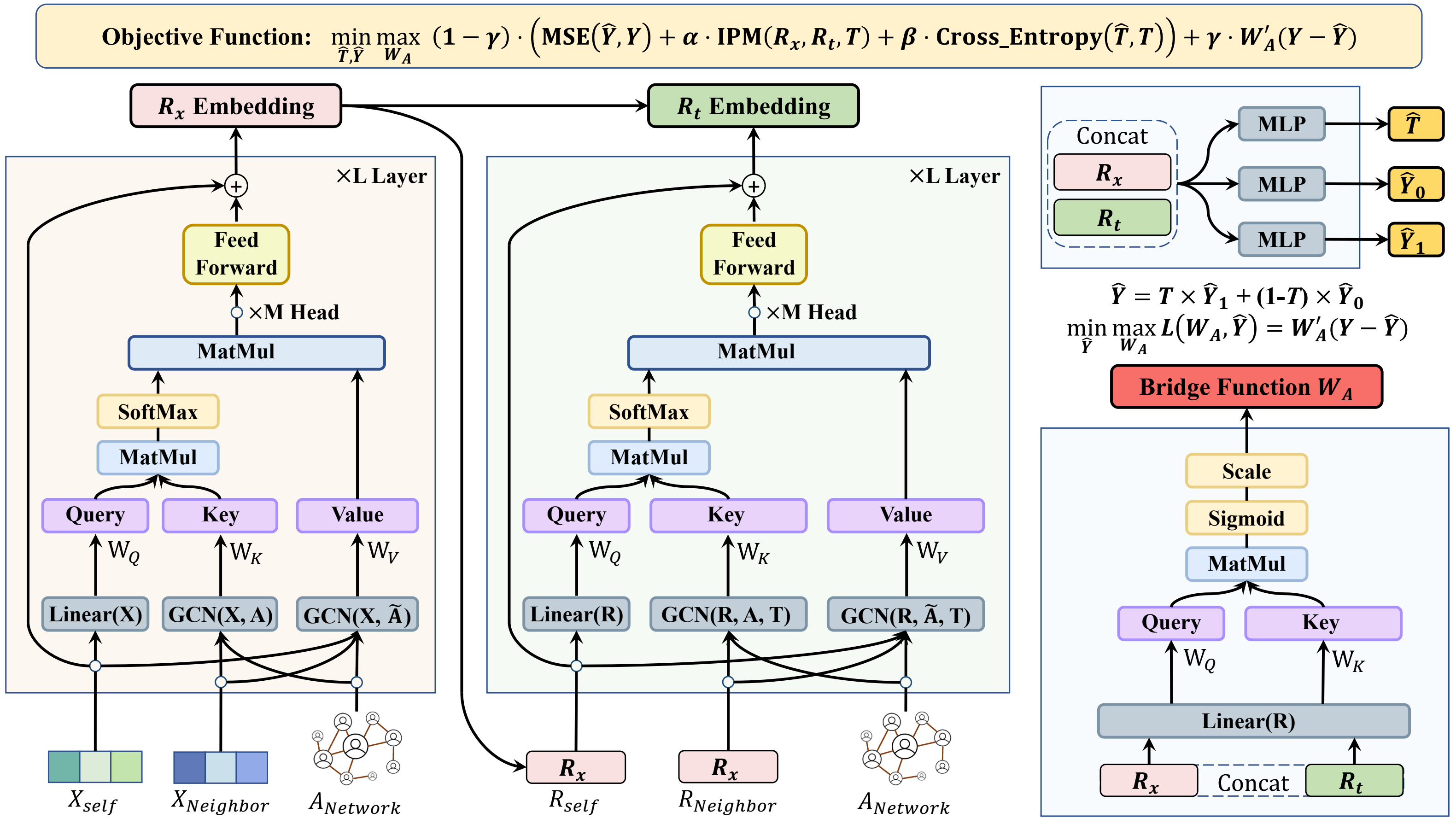
$$\begin{aligned}\tau_{\text{ITE}}(\boldsymbol{x}_i, \boldsymbol{t}_{\mathcal{P}}) &= \mathbb{E}[y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 1, \boldsymbol{t}_{\mathcal{P}}) \mid \boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}] - \mathbb{E}[y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 0, \boldsymbol{0}_{\mathcal{P}}) \mid \boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}] \\ &= \mathbb{E}[y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 1, \boldsymbol{t}_{\mathcal{P}}) \mid \boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, t, \boldsymbol{t}_{\mathcal{P}}] - \mathbb{E}[y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 0, \boldsymbol{0}_{\mathcal{P}}) \mid \boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 0, \boldsymbol{0}_{\mathcal{P}}] \qquad (1) \\ &= \mathbb{E}[y_{1,\boldsymbol{t}_{\mathcal{P}}} \mid \boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, t, \boldsymbol{t}_{\mathcal{P}}] - \mathbb{E}[y_{0,\boldsymbol{0}_{\mathcal{P}}} \mid \boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, 0, \boldsymbol{0}_{\mathcal{P}}], \qquad (2)\end{aligned}$$

where $y_{t,\boldsymbol{t}_{\mathcal{P}}}$ is the observed outcome when the unit and its peers have features $\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}$ and receive the treatment pair $(t, \boldsymbol{t}_{\mathcal{P}})$. Eq. (1) holds under the Uncounfoundedness, i.e., $y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, t, \boldsymbol{t}_{\mathcal{P}}) \perp\!\!\!\perp (t, \boldsymbol{t}_{\mathcal{P}}) \mid (\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i})$. Eq. (2) holds under the Consistency Assumption, i.e., $y_{t,\boldsymbol{t}_{\mathcal{P}}} = y(\boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, t, \boldsymbol{t}_{\mathcal{P}})$. Theorem 1 holds for any treatment pair $(t, \boldsymbol{t}_{\mathcal{P}})$ under Positivity Assumption, i.e., $\mathbb{P}(t_i = t, \boldsymbol{t}_{\mathcal{P}_i} = \boldsymbol{t}_{\mathcal{P}}) > 0$. □

**Assumption 4** (Representation). *The treatment vector $\boldsymbol{t}_{\mathcal{P}_i}$ of peer nodes can be captured by a peer treatment function $g_t$, and the confounder vector $\boldsymbol{x}_{\mathcal{P}_i}$ by a peer confounder function $g_x$.*

**Proposition 1** (Interference). *If the unknown interference graph $\boldsymbol{E}$ is latent in full graph information $\{\boldsymbol{x}, \boldsymbol{t}, \boldsymbol{A}\}$, then, the outcome $\mathbb{E}[y_{t,\boldsymbol{t}_{\mathcal{P}}} \mid \boldsymbol{x}_i, \boldsymbol{x}_{\mathcal{P}_i}, t, \boldsymbol{t}_{\mathcal{P}}]$ is identified.*

# Causality-based Graph Transformer (CauGramer)



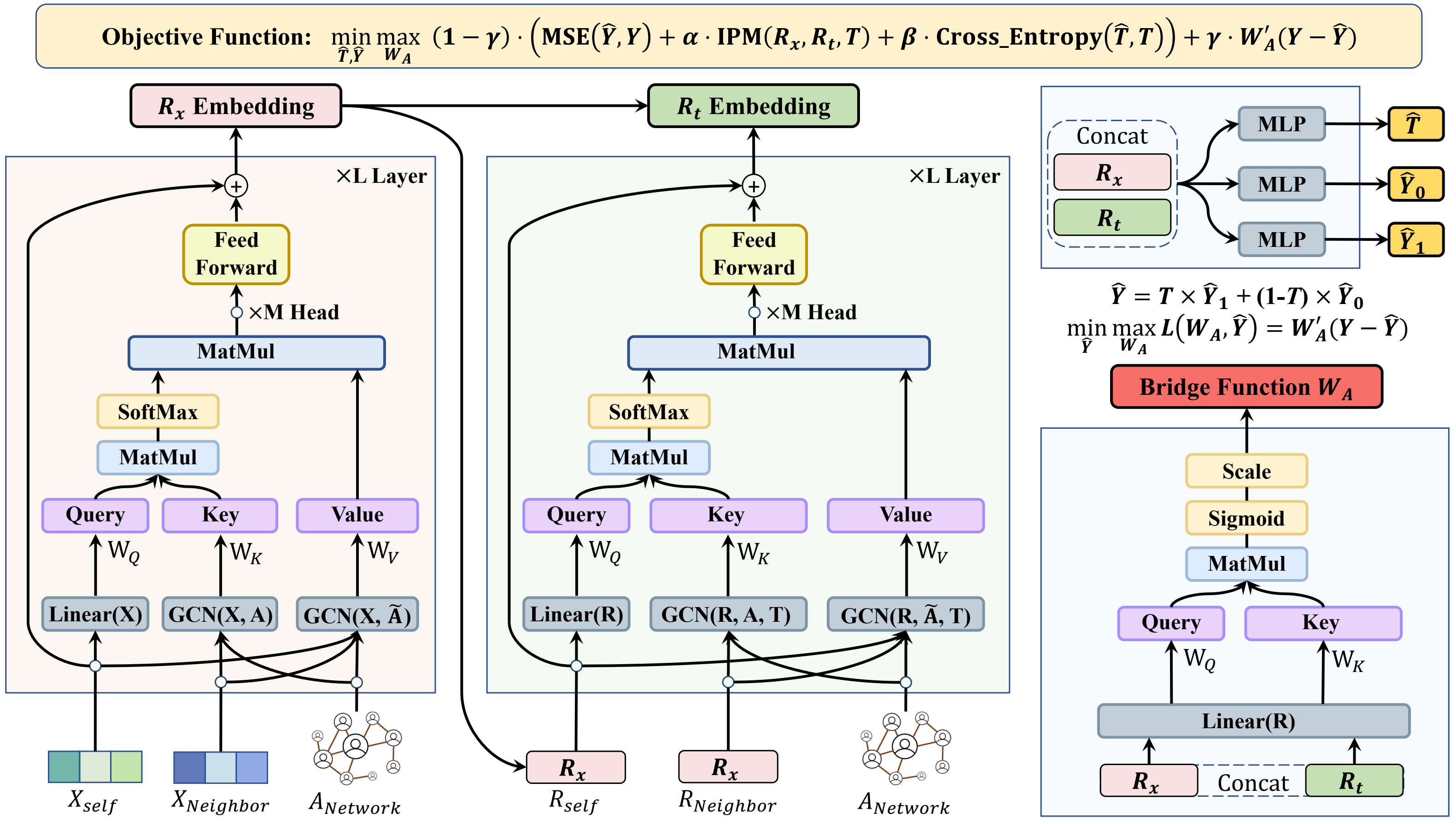

## Advance: Graph Transformer

This paper designs a causality-based graph transformer that models network interference by constructing linear queries from individual features, graph convolutional keys from peer features, and combined values. This approach expands the receptive field of the graph neural network while capturing complex interference patterns.

$$\boldsymbol{r}_x^{(h+1)} = \mathrm{Attention}_x^{(h)} \cdot \mathrm{V}_x^{(h)} = \mathrm{Softmax}\left(\frac{\mathrm{Q}_x^{(h)} \cdot \mathrm{K}_x^{(h)\prime}}{\sqrt{d}}\right) \cdot \mathrm{V}_x^{(h)}, \quad (4)$$

$$\mathrm{Q}_x^{(h)} = \mathrm{Linear}_x^{(h)}(\boldsymbol{r}_x^{(h)})W_Q^{(h)\prime}, \quad \mathrm{K}_x^{(h)} = \mathrm{GCN}_{xk}^{(h)}(\boldsymbol{r}_x^{(h)}, \boldsymbol{A})W_K^{(h)\prime}, \quad \mathrm{V}_x^{(h)} = \mathrm{GCN}_{xv}^{(h)}(\boldsymbol{r}_x^{(h)}, \tilde{\boldsymbol{A}})W_V^{(h)\prime}. \quad (5)$$

# Causality-based Graph Transformer (CauGramer)




## Advance: Joint Balancing

Traditional algorithms assume that peer-treatments would not influent self-treatment, only address the confounding bias in main effect estimation.

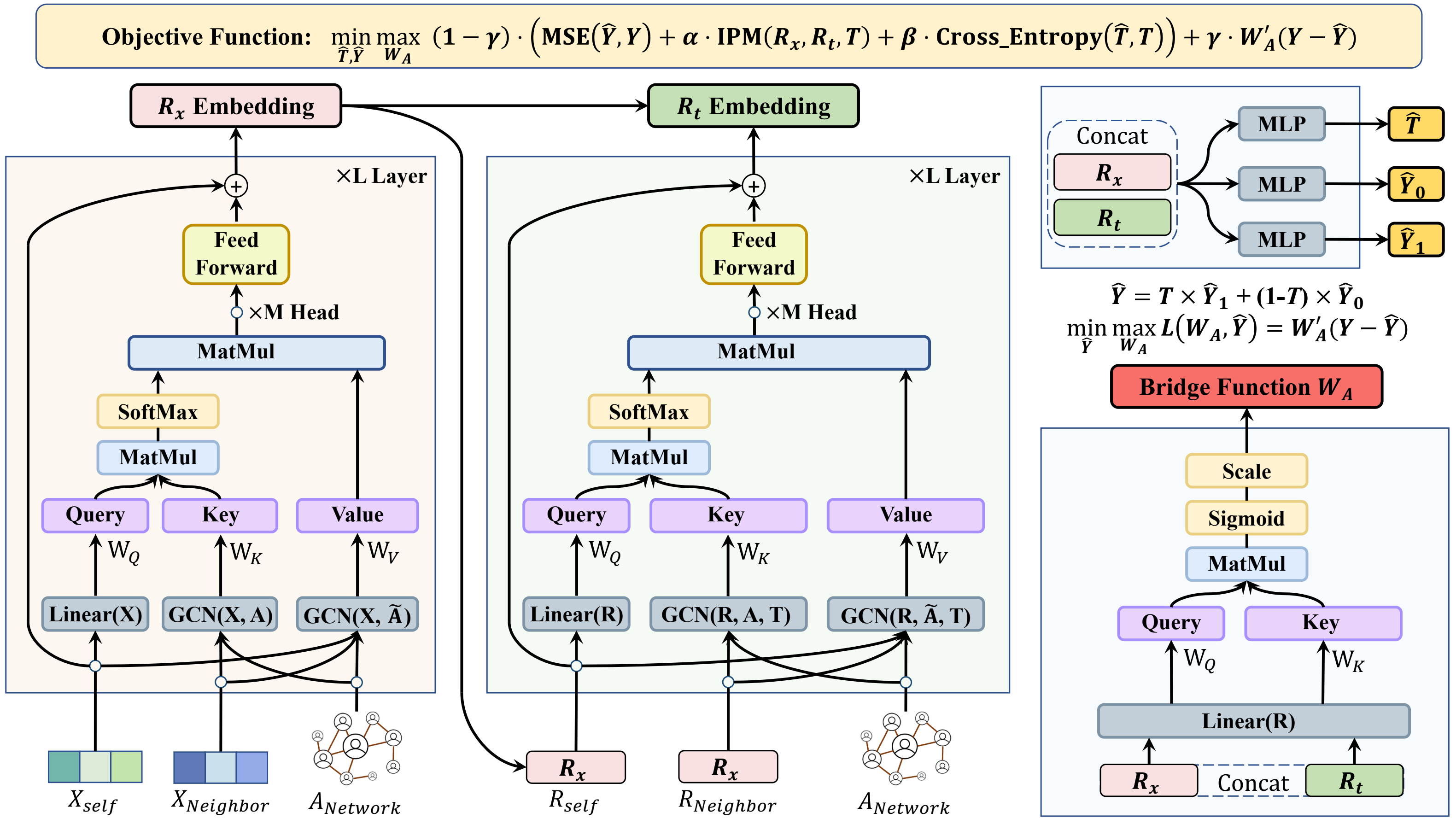
In this work, we proposes a joint representation balancing, which relaxes the previous treatment-independence assumption.

$$\mathrm{IPM}(\boldsymbol{r}, \boldsymbol{t}) = \mathrm{Wass}(\{\boldsymbol{r}_{i:t_i=0}\}, \{\boldsymbol{r}_{j:t_j=1}\}), \tag{8}$$

$$\mathrm{Cross\_Entropy}(\hat{\boldsymbol{t}}, \boldsymbol{t}) = -\frac{1}{N}\sum_{i=1}^{N}\left[t_i \log\left(p_i\right) + (1 - t_i)\log\left(1 - p_i\right)\right], \tag{9}$$

$$\mathcal{L}_y = \mathrm{MSE}(\hat{\boldsymbol{y}}, \boldsymbol{y}) + \alpha \cdot \mathrm{IPM}(\boldsymbol{r}, \boldsymbol{t}) + \beta \cdot \mathrm{Cross\_Entropy}(\hat{\boldsymbol{t}}, \boldsymbol{t}), \tag{10}$$

# Causality-based Graph Transformer (CauGramer)




## Advance: Minimax Constraint

We refine the potential outcome prediction model using a minimax moment constraint, enabling it to correct for confounding bias through bridge function, even in the presence of unmeasured confounders.

$$y^* = \arg\min_{\hat{y}} \max_{q\in\mathbb{Q}} \mathbb{E}[(y-\hat{y})q(r,t)], \quad q(r,t) = \mathrm{Sigmoid}\left(\mathrm{Q}\cdot \mathrm{K}'\right), \tag{11}$$

# Experiment Results

## Datasets.

Following previous works, we use pseudo-real datasets from BlogCatalog (BC) and Flickr, where the features (x) and social networks (A) are real, while treatments (t), outcomes (y), and interference (E) are simulated.

## Evaluation.

We use the Average Absolute Error ($\epsilon_{AVG}$) on AME, APE, and ATE as evaluation metric

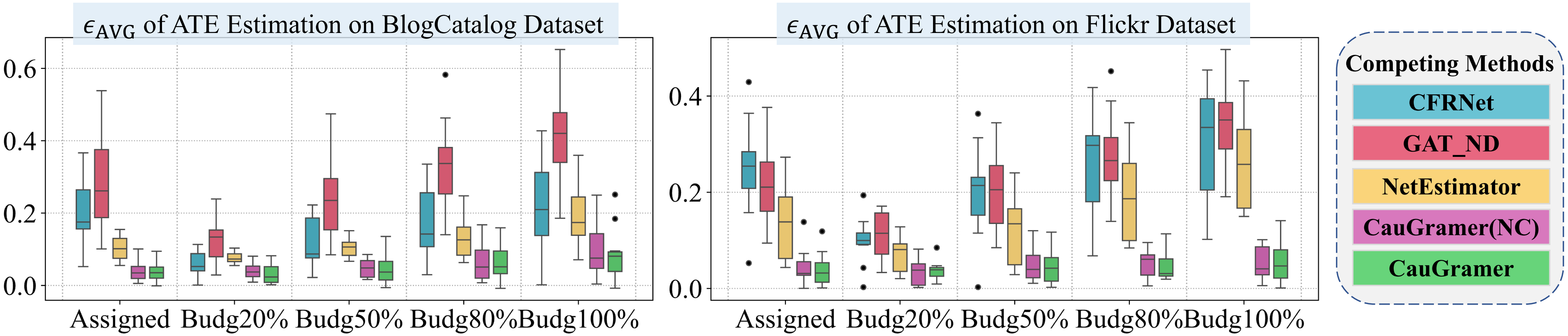
We use ($\sqrt{\epsilon_{PEHE}}$) on IME, IPE, and ITE as evaluation metric.

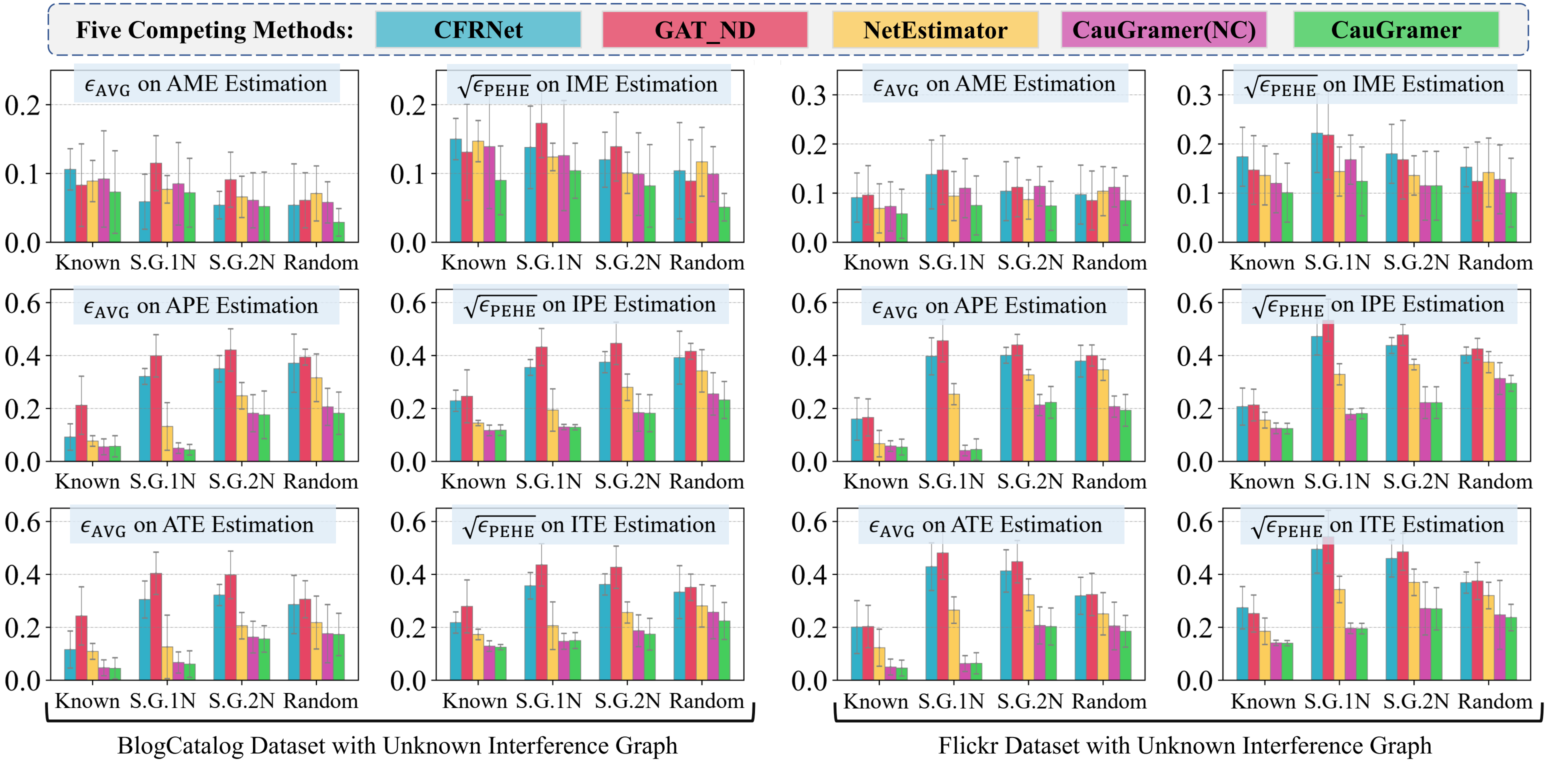
Table 2: Results of Constant Treatment Effects Estimation on BlogCatalog (BC) and Flickr Datasets.

| **BC** | Effects | CFRNet | DRLearner | NetDeconf | G-HSIC | SPNet | CAL | Graphormer | RRNet | NetEst | CauGramer |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\epsilon_{AVE}$ | AME | $\underline{0.058}_{\pm0.03}$ | $0.210_{\pm0.03}$ | $0.075_{\pm0.03}$ | $0.076_{\pm0.04}$ | $0.066_{\pm0.04}$ | $0.083_{\pm0.03}$ | $0.086_{\pm0.06}$ | $0.105_{\pm0.05}$ | $0.076_{\pm0.02}$ | $\mathbf{0.054}_{\pm0.03}$ |
| | APE | $0.117_{\pm0.06}$ | $0.207_{\pm0.02}$ | $0.351_{\pm0.09}$ | $0.387_{\pm0.02}$ | $0.223_{\pm0.10}$ | $0.370_{\pm0.07}$ | $0.436_{\pm0.03}$ | $0.229_{\pm0.06}$ | $\underline{0.078}_{\pm0.02}$ | $\mathbf{0.034}_{\pm0.03}$ |
| | ATE | $0.123_{\pm0.07}$ | $\underline{0.064}_{\pm0.05}$ | $0.337_{\pm0.09}$ | $0.351_{\pm0.06}$ | $0.203_{\pm0.10}$ | $0.355_{\pm0.08}$ | $0.349_{\pm0.07}$ | $0.296_{\pm0.06}$ | $0.065_{\pm0.03}$ | $\mathbf{0.039}_{\pm0.04}$ |
| $\sqrt{\epsilon_{PEHE}}$ | IME | $\underline{0.096}_{\pm0.03}$ | $0.545_{\pm0.01}$ | $0.119_{\pm0.05}$ | $0.132_{\pm0.01}$ | $0.100_{\pm0.05}$ | $0.131_{\pm0.03}$ | $0.211_{\pm0.09}$ | $0.152_{\pm0.07}$ | $0.099_{\pm0.03}$ | $\mathbf{0.075}_{\pm0.07}$ |
| | IPE | $0.122_{\pm0.06}$ | $0.220_{\pm0.02}$ | $0.365_{\pm0.09}$ | $0.410_{\pm0.03}$ | $0.230_{\pm0.10}$ | $0.384_{\pm0.07}$ | $0.457_{\pm0.03}$ | $0.238_{\pm0.06}$ | $\underline{0.092}_{\pm0.01}$ | $\mathbf{0.044}_{\pm0.02}$ |
| | ITE | $0.147_{\pm0.07}$ | $0.514_{\pm0.01}$ | $0.351_{\pm0.09}$ | $0.384_{\pm0.06}$ | $0.213_{\pm0.09}$ | $0.370_{\pm0.08}$ | $0.429_{\pm0.04}$ | $0.311_{\pm0.06}$ | $\underline{0.117}_{\pm0.01}$ | $\mathbf{0.063}_{\pm0.04}$ |
| **Flickr** | Effects | CFRNet | DRLearner | NetDeconf | G-HSIC | SPNet | CAL | Graphormer | RRNet | NetEst | CauGramer |
| $\epsilon_{AVE}$ | AME | $0.066_{\pm0.04}$ | $0.110_{\pm0.05}$ | $0.088_{\pm0.03}$ | $0.096_{\pm0.03}$ | $0.054_{\pm0.03}$ | $0.090_{\pm0.05}$ | $\underline{0.030}_{\pm0.03}$ | $0.160_{\pm0.03}$ | $0.063_{\pm0.04}$ | $\mathbf{0.028}_{\pm0.03}$ |
| | APE | $0.115_{\pm0.04}$ | $0.211_{\pm0.03}$ | $0.345_{\pm0.06}$ | $0.354_{\pm0.04}$ | $0.121_{\pm0.05}$ | $0.302_{\pm0.04}$ | $0.409_{\pm0.03}$ | $0.268_{\pm0.09}$ | $\underline{0.055}_{\pm0.03}$ | $\mathbf{0.019}_{\pm0.01}$ |
| | ATE | $0.144_{\pm0.06}$ | $0.117_{\pm0.07}$ | $0.351_{\pm0.05}$ | $0.300_{\pm0.09}$ | $0.131_{\pm0.05}$ | $0.310_{\pm0.06}$ | $0.382_{\pm0.04}$ | $0.374_{\pm0.11}$ | $\underline{0.073}_{\pm0.05}$ | $\mathbf{0.032}_{\pm0.02}$ |
| $\sqrt{\epsilon_{PEHE}}$ | IME | $0.119_{\pm0.04}$ | $0.517_{\pm0.01}$ | $0.134_{\pm0.05}$ | $0.129_{\pm0.02}$ | $\underline{0.079}_{\pm0.05}$ | $0.137_{\pm0.07}$ | $0.086_{\pm0.03}$ | $0.229_{\pm0.04}$ | $0.095_{\pm0.05}$ | $\mathbf{0.047}_{\pm0.04}$ |
| | IPE | $0.128_{\pm0.04}$ | $0.240_{\pm0.03}$ | $0.382_{\pm0.07}$ | $0.405_{\pm0.04}$ | $0.131_{\pm0.05}$ | $0.332_{\pm0.05}$ | $0.459_{\pm0.03}$ | $0.294_{\pm0.10}$ | $\underline{0.070}_{\pm0.03}$ | $\mathbf{0.032}_{\pm0.01}$ |
| | ITE | $0.177_{\pm0.05}$ | $0.531_{\pm0.02}$ | $0.380_{\pm0.05}$ | $0.363_{\pm0.09}$ | $0.142_{\pm0.05}$ | $0.334_{\pm0.06}$ | $0.443_{\pm0.04}$ | $0.405_{\pm0.12}$ | $\underline{0.105}_{\pm0.05}$ | $\mathbf{0.051}_{\pm0.02}$ |

Table 3: Results of Heterogeneous Treatment Effects Estimation with/without Unconfoundedness on BlogCatalog (BC) and Flickr Datasets. The best is **boldface** while the second best is underlined.

| **BC** | Effects | CFRNet | DRLearner | GDML | SPNet | NetEst | CEVAE | CNE | UNITE | C.G.(UC) | C.G.(NC) | CauGramer |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\epsilon_{AVE}$ | AME | $0.106_{\pm0.03}$ | $0.196_{\pm0.07}$ | $0.166_{\pm0.06}$ | $0.083_{\pm0.06}$ | $0.089_{\pm0.03}$ | $0.081_{\pm0.03}$ | $0.186_{\pm0.04}$ | $\mathbf{0.069}_{\pm0.00}$ | $0.109_{\pm0.07}$ | $0.092_{\pm0.07}$ | $\underline{0.073}_{\pm0.06}$ |
| | APE | $0.092_{\pm0.05}$ | $0.183_{\pm0.03}$ | $0.371_{\pm0.06}$ | $0.212_{\pm0.11}$ | $0.077_{\pm0.02}$ | $0.403_{\pm0.01}$ | $0.519_{\pm0.05}$ | - | $0.067_{\pm0.04}$ | $\mathbf{0.055}_{\pm0.03}$ | $\underline{0.057}_{\pm0.04}$ |
| | ATE | $0.116_{\pm0.07}$ | $0.099_{\pm0.07}$ | $0.537_{\pm0.09}$ | $0.243_{\pm0.11}$ | $0.109_{\pm0.03}$ | $0.349_{\pm0.04}$ | $1.127_{\pm0.06}$ | - | $0.077_{\pm0.05}$ | $\underline{0.047}_{\pm0.03}$ | $\mathbf{0.045}_{\pm0.04}$ |
| $\sqrt{\epsilon_{PEHE}}$ | IME | $0.150_{\pm0.03}$ | $0.544_{\pm0.03}$ | $0.265_{\pm0.08}$ | $0.131_{\pm0.07}$ | $0.147_{\pm0.03}$ | $\underline{0.120}_{\pm0.03}$ | $0.288_{\pm0.05}$ | $0.194_{\pm0.00}$ | $0.159_{\pm0.10}$ | $0.139_{\pm0.09}$ | $\mathbf{0.090}_{\pm0.05}$ |
| | IPE | $0.229_{\pm0.04}$ | $0.216_{\pm0.03}$ | $0.411_{\pm0.06}$ | $0.246_{\pm0.10}$ | $0.145_{\pm0.01}$ | $0.439_{\pm0.01}$ | $0.552_{\pm0.05}$ | - | $0.125_{\pm0.02}$ | $\mathbf{0.117}_{\pm0.02}$ | $\underline{0.118}_{\pm0.02}$ |
| | ITE | $0.218_{\pm0.04}$ | $0.529_{\pm0.02}$ | $0.607_{\pm0.10}$ | $0.279_{\pm0.10}$ | $0.173_{\pm0.02}$ | $0.398_{\pm0.03}$ | $0.610_{\pm0.06}$ | - | $0.149_{\pm0.03}$ | $\underline{0.129}_{\pm0.02}$ | $\mathbf{0.125}_{\pm0.01}$ |
| **Flickr** | Effects | CFRNet | DRLearner | GDML | SPNet | NetEst | CEVAE | CNE | UNITE | C.G.(UC) | C.G.(NC) | CauGramer |
| $\epsilon_{AVE}$ | AME | $0.091_{\pm0.05}$ | $0.135_{\pm0.08}$ | $0.239_{\pm0.08}$ | $0.096_{\pm0.06}$ | $0.069_{\pm0.05}$ | $0.063_{\pm0.04}$ | $0.168_{\pm0.05}$ | $\mathbf{0.043}_{\pm0.00}$ | $0.091_{\pm0.07}$ | $0.073_{\pm0.05}$ | $\underline{0.058}_{\pm0.05}$ |
| | APE | $0.160_{\pm0.08}$ | $0.216_{\pm0.02}$ | $0.381_{\pm0.02}$ | $0.166_{\pm0.07}$ | $0.067_{\pm0.05}$ | $0.432_{\pm0.05}$ | $0.562_{\pm0.06}$ | - | $0.067_{\pm0.02}$ | $\underline{0.038}_{\pm0.02}$ | $\mathbf{0.030}_{\pm0.03}$ |
| | ATE | $0.201_{\pm0.10}$ | $0.131_{\pm0.07}$ | $0.620_{\pm0.08}$ | $0.203_{\pm0.08}$ | $0.123_{\pm0.07}$ | $0.389_{\pm0.05}$ | $1.055_{\pm0.02}$ | - | $0.056_{\pm0.04}$ | $\underline{0.032}_{\pm0.03}$ | $\mathbf{0.025}_{\pm0.03}$ |
| $\sqrt{\epsilon_{PEHE}}$ | IME | $0.174_{\pm0.06}$ | $0.529_{\pm0.02}$ | $0.359_{\pm0.10}$ | $0.147_{\pm0.07}$ | $0.136_{\pm0.06}$ | $0.139_{\pm0.04}$ | $0.261_{\pm0.06}$ | $0.212_{\pm0.00}$ | $0.141_{\pm0.08}$ | $\underline{0.107}_{\pm0.06}$ | $\mathbf{0.096}_{\pm0.06}$ |
| | IPE | $0.207_{\pm0.07}$ | $0.264_{\pm0.02}$ | $0.448_{\pm0.02}$ | $0.213_{\pm0.06}$ | $0.156_{\pm0.03}$ | $0.505_{\pm0.05}$ | $0.637_{\pm0.06}$ | - | $0.128_{\pm0.01}$ | $\underline{0.112}_{\pm0.02}$ | $\mathbf{0.111}_{\pm0.01}$ |
| | ITE | $0.274_{\pm0.08}$ | $0.547_{\pm0.02}$ | $0.716_{\pm0.09}$ | $0.252_{\pm0.07}$ | $0.185_{\pm0.05}$ | $0.478_{\pm0.04}$ | $0.669_{\pm0.07}$ | - | $0.150_{\pm0.02}$ | $\underline{0.127}_{\pm0.01}$ | $\mathbf{0.125}_{\pm0.01}$ |




**Conclusion:** By constructing linear queries from individual features, graph convolutional keys from peer features, and combined values to model network interference, CauGramer expands the receptive field of the graph neural network while capturing complex interference patterns. Experiments on two widely used benchmark datasets demonstrate that the proposed CauGramer outperforms existing methods in network causal effect estimation.

# Thanks

## Acknowledgement

This work was supported by the National Natural Science Foundation of China (62441605, 62441617, 62376243, 62037001), and the Starry Night Science Fund at Shanghai Institute for Advanced Study (Zhejiang University).